

# Math-Net.Ru

All Russian mathematical portal

A. I. Kaplinskiĭ, A. I. Propoĭ, Nonlocal optimization methods that use potential theory, *Avtomat. i Telemekh.*, 1993, Issue 7, 55–65

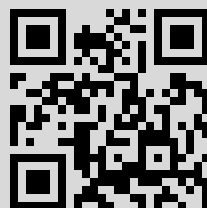
Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use

<http://www.mathnet.ru/eng/agreement>

Download details:

IP: 67.81.248.28

August 4, 2022, 18:30:58



УДК 517.977

© 1993 г. А.И. КАПЛИНСКИЙ, канд. техн. наук  
(Воронежский государственный университет),

А.И. ПРОПОЙ, д-р техн. наук  
(Институт системного анализа РАН, Москва)

## МЕТОДЫ НЕЛОКАЛЬНОЙ ОПТИМИЗАЦИИ, ИСПОЛЬЗУЮЩИЕ ТЕОРИЮ ПОТЕНЦИАЛА

Рассматривается подход, основанный на замене задачи минимизации функции многих переменных на эквивалентную рандомизированную задачу. В результате вариации функционала рандомизированной задачи показано, что исходная задача сводится к максимизации потенциальной функции, являющейся решением некоторой краевой задачи математической физики. Рассмотрены методы первого и второго порядка (аналоги градиентного спуска и метода Ньютона применительно к потенциальной функции), которые для своей реализации требуют лишь значения исходной функции. На основе анализа свойств потенциальной функции выявлены основные структурные свойства нелокального поиска: неустойчивая составляющая в неперспективной области поиска, операции отражения и растяжения соответственно в неперспективном и перспективном направлении.

### 1. Введение

Методы решения экстремальных задач к настоящему времени достаточно хорошо разработаны. Однако эти методы в основном ориентированы на решение "хороших" задач (выпуклых, гладких, детерминированных) и основаны на использовании локальной информации (типа "градиент в точке"). Принципиальная черта "плохих" экстремальных задач (невыпуклых, негладких, недетерминированных) состоит в том, что они требуют нелокальной информации в процессе их решения.

Общим путем введения нелокальности является рандомизация исходной задачи. Для задач нелинейного программирования это было сделано в [1–4]. Методы решения таких задач основывались на стохастической модификации (обычного) градиента и приводили к вероятностным итеративным методам их решения (см., например, [4–6]).

Цель настоящей работы – рассмотреть альтернативный (двойственный) подход к понятию градиента в негладкой оптимизации и следующие из него методы. Это направление имеет глубокие корни с методами адаптации и обучения [7], теорией потенциальных функций [8], математической теорией систем [9] и сформировалось сравнительно недавно [10–17].

В работе в систематизированном виде изложены основы теории этого направления. Рассмотрена вариация первого порядка функционала рандомизированной задачи. Даны условия, которым должно удовлетворять усредненное векторное поле, приводящее к улучшению решения. В результате его разложения на потенциальную и "бездивергентную" составляющую показано, что исходную задачу можно свести к максимизации потенциальной

функции, являющейся решением краевой задачи математической физики. Рассмотрены методы первого и второго порядка (аналоги градиентного спуска и метода Ньютона применительно к потенциальной функции), для реализации которых требуется вычисление лишь значений исходной функции. Показано, каким способом можно учитывать априорные сведения о задаче (учет дифференциальных свойств целевой функции, предположений о локализации решения). На основе анализа свойств потенциальной функции выявлены основные структурные свойства нелокального поиска: неустойчивая компонента в неперспективной области, операции отражения и растяжения пространства в неперспективном и перспективном направлениях поиска.

## 2. Рандомизация задачи

Рассматривается задача безусловной оптимизации

$$(1) \quad f(x) \rightarrow \min,$$

где функция  $f: R^n \rightarrow R$ .

При построении методов решения задачи (1) будем ориентироваться на "плохую" функцию  $f$ , предполагая, что в общем случае измеримы лишь значения  $f(x)$ . Однако при этом хотелось бы иметь технологию построения таких методов решения (1), которые учитывали бы уровень информированности о решаемой задаче (знание производных, области локализации решения и т.д.).

Для этой цели переформулируем задачу (1), вводя распределение  $x$ , т.е. осуществляя рандомизацию (1):

$$(2) \quad F(X) = E[f(X)] \rightarrow \min_{X \in \{X\}}.$$

Здесь  $\{X\}$  – множество допустимых случайных векторов со значениями в  $R^n$ .

Задача (2) может быть также переписана в виде

$$(3) \quad \int_{R^n} f(x)p(x)dx \rightarrow \min_{p(x)},$$

где  $p(x)$  – плотность распределения вероятностей случайного вектора  $X$ , понимаемая, вообще говоря, в обобщенном смысле.

Возможность замены задачи (1) на рандомизированную (2) основана на следующих утверждениях.

**Предложение 1.** Если множество случайных векторов  $\{X\}$  включает в себя множество дискретно распределенных векторов, т.е. множество допустимых распределений  $p(x)$  включает  $\delta$ -функции, то

$$(4) \quad \min_{x \in R^n} f(x) = \min_{X \in \{X\}} E[f(X)].$$

**Предложение 2.** Пусть  $X^*$  – случайный вектор с распределением  $p^*(x)$  – решение (2). Тогда любая точка  $x^* \in R^n$ , для которой  $f(x)$  непрерывна и при

$$\|x - x^*\| \leq \varepsilon \quad \int_{\|x - x^*\| \leq \varepsilon} p^*(x)dx > 0, \text{ будет решением задачи (1).}$$

Доказательство этих предложений достаточно очевидно и может быть найдено, например, в [11].

Рассмотрим задачу нелинейного программирования

$$(5) \quad \omega = \{\min f(x) | f_j(x) \leq 0, j = 1, \dots, m, x \in R^n\}$$

и ее рандомизацию [2]:

$$(6) \quad \bar{\omega} = \{\min \int_{R^n} f(x)p(x)dx | \int_{R^n} f_j(x)p(x)dx \leq 0, j = 1, \dots, m\}.$$

В этом случае, в отличие от (4),  $\bar{\omega} \leq \omega$ . Это происходит потому, что множество допустимых решений (6), вообще говоря, шире, чем в (5) (ограничения в (6) могут выполняться лишь в среднем). Если же (5) – задача выпуклого программирования, то имеют место утверждения относительно эквивалентности задач (5) и (6), аналогичные предложениям 1 и 2.

Вариантом задачи (2) является задача

$$(7) \quad \bar{\omega}_{\Omega} = \{\min_{\Omega} \int f(x)p(x)dx | x \in \Omega\},$$

где  $\Omega$  – некоторое множество в  $R^n$ , причем  $\int_{\Omega} p(x)dx = 1$  (например,  $\Omega$  может характеризовать априорную информацию о локализации решения задачи). В этом случае также имеют место утверждения об эквивалентности задач, даже если множество  $\Omega$  невыпукло (так как здесь рандомизируется целевая функция, а не ограничения).

### 3. Вариация рандомизированного функционала

Для решения задачи (2) будем рассматривать итеративные процедуры, имеющие в пространстве случайных векторов обычный вид

$$(8) \quad X_{N+1} = X_N + \varepsilon_N Y_N.$$

Здесь  $Y_N$  – случайное направление изменения  $X_N$  с совместной плотностью распределения  $p_N(x, y)$ ,  $\varepsilon_N$  – длина шага вдоль  $Y_N$ .

Для нахождения  $Y_N$  определим  $\delta_Y F(X)$  – производную функционала (2) в состоянии  $X$  по направлению  $Y$ .

По определению

$$(9) \quad \delta_Y F(X) = \lim_{\varepsilon \rightarrow 0} \frac{F(X + \varepsilon Y) - F(X)}{\varepsilon} = \frac{d}{d\varepsilon} [F(X + \varepsilon Y)]_{\varepsilon=0} = \frac{d}{d\varepsilon} [F(X_{\varepsilon})]_{\varepsilon=0} =$$

$$= \left[ \frac{d}{d\varepsilon} \int_{R^n} f(x) p_{\varepsilon}(x) dx \right]_{\varepsilon=0},$$

где плотность распределения  $p_{\varepsilon}(x)$  случайного вектора  $X_{\varepsilon} = X + \varepsilon Y$  может быть выражена через совместную плотность

$$(10) \quad p_{\varepsilon}(x) = \int_{R^n} p(x - \varepsilon y, y) dy.$$

Предполагая выполненными условия дифференцируемости интегралов по параметру (они всегда могут быть обеспечены соответствующим выбором распределений), получим

$$(11) \quad \delta_Y F(X) = \left[ \frac{d}{d\varepsilon} \int_{R^n} f(x) \int_{R^n} p(x - \varepsilon y, y) dy dx \right]_{\varepsilon=0} = \int_{R^n} \frac{d}{d\varepsilon} [A_{\varepsilon} f(y)]_{\varepsilon=0} dy.$$

Здесь

$$(12) \quad A_{\varepsilon} f = \int_{R^n} f(x) p(x - \varepsilon y, y) dx$$

– оператор псевдодифференциального типа [18].

Для вычисления (11) воспользуемся известным соотношением векторного анализа

$$\left[ \frac{d}{d\varepsilon} p(x - \varepsilon y, y) \right]_{\varepsilon=0} = - \left( \frac{\partial p(x, y)}{\partial x}, y \right) = - \operatorname{div}_x [p(x, y) y],$$

где  $(, )$  – скалярное произведение в  $R^n$ ,  $\operatorname{div} f = \sum_{i=1}^n \frac{\partial f}{\partial x_i}$ . В результате

$$\begin{aligned}\delta_Y F(X) &= - \int_{R^n} \int_{R^n} f(x) \left( \frac{\partial p(x, y)}{\partial x}, y \right) dy dx = \\ &= - \int_{R^n} \int_{R^n} f(x) \operatorname{div}_x [p(x, y) y] dy dx = - \int_{R^n} f(x) \operatorname{div}_x \left[ \int_{R^n} p(x, y) y dy \right] dx.\end{aligned}$$

Используя условную плотность вероятности  $p(x, y) = p(x)p(y|x)$  и обозначая

$$(13) \quad \bar{y}(x) = \int_{R^n} y p(y|x) dy = E[Y|X = x],$$

перепишем последнее выражение в виде

$$\delta_Y F(X) = - \int_{R^n} f(x) \operatorname{div} [p(x) \bar{y}(x)] dx.$$

Дифференцируя по  $\varepsilon$  тождество  $\int_{R^n} p_\varepsilon(x) dx \equiv 1$ , получим

$$(14) \quad \int_{R^n} \operatorname{div} [p(x) \bar{y}(x)] dx = 0.$$

Окончательно получим для производной  $F$  по направлению  $Y$  выражение

$$(15) \quad \delta_Y F(X) = - \int_{R^n} [f(x) - c] \operatorname{div} [p(x) \bar{y}(x)] dx,$$

где  $c$  – произвольная константа.

Направление  $Y$ , которое приводит к уменьшению  $F$ , должно удовлетворять неравенству  $\delta_Y F(X) \leq 0$ . В частности, оно будет выполняться, если

$$\operatorname{div} [p(x) \bar{y}(x)] = [f(x) - c] p_U(x), \quad p_U(x) \geq 0.$$

Подробно выполнение этого неравенства обсуждается в следующем разделе.

Пусть функция  $f(x)$  имеет непрерывные производные первого порядка. Тогда  $\delta_Y F(X)$  может быть представлена в виде

$$\begin{aligned}\delta_Y F(X) &= \left[ \frac{d}{d\varepsilon} \int_{R^n} \int_{R^n} f(x + \varepsilon y) p(x, y) dx dy \right]_{\varepsilon=0} = \\ &= \int_{R^n} \int_{R^n} \left( \frac{\partial f(x)}{\partial x}, y \right) p(x, y) dx dy = \int_{R^n} \left( \frac{\partial f(x)}{\partial x}, \bar{y}(x) \right) p(x) dx,\end{aligned}$$

где  $\bar{y}(x)$  определено в (13).

Таким образом, производная по направлению  $\delta_Y F(X)$  имеет два эквивалентных представления:

$$(16) \quad \delta_Y F(X) = \int_{R^n} \left( \frac{\partial f(x)}{\partial x}, \bar{y}(x) \right) p(x) dx = - \int_{R^n} (f(x) - c) \operatorname{div} [p(x) \bar{y}(x)] dx.$$

Первое представление в (16) ведет к известным вероятностным итеративным методам оптимизации (типа стохастической аппроксимации) [7], в то время как второе представление в (16) является основой новых методов, которые будут рассмотрены в следующем разделе.

#### 4. Понятие градиента в негладкой оптимизации

Определим теперь направление  $Y$ , которое максимизирует  $\delta_Y F(X)$  (при некоторых естественных ограничениях на  $Y$ , чтобы избежать бесконечных решений). Такое направление будем называть градиентным. Для этой цели воспользуемся обобщенным неравенством Шварца с весовой функцией  $p_U(x)$ :

$$(17) \quad |\delta_Y F(X)| = \left( \int_{R^n} (f(x) - c)^2 p_U(x) dx \right)^{1/2} \left( \int_{R^n} \left\{ \frac{\operatorname{div}[p(x)\bar{y}(x)]}{p_U(x)} \right\}^2 p_U(x) dx \right)^{1/2},$$

где  $p_U(x) > 0$  может рассматриваться как плотность распределения вероятности некоторого случайного вектора  $U$ . Соответствующий выбор  $p_U(x)$  позволяет обеспечить существование интегралов в (17).

Из (17) следует, что максимизация  $\delta_Y F(X)$  по  $Y$  приводит к уравнению

$$\frac{\operatorname{div}[p(x)\bar{y}(x)]}{p_U(x)} = \lambda(f(x) - c), \lambda > 0$$

или

$$(18) \quad \operatorname{div}[p(x)\bar{y}(x)] = \lambda(f(x) - c)p_U(x),$$

где теперь вследствие (14) постоянная  $c$  определяется равенством

$$(19) \quad c = \int_{R^n} f(x)p_U(x)dx.$$

Если ограничить "длину" вектора  $Y$ :

$$(20) \quad \int_{R^n} \left\{ \frac{\operatorname{div}[p(x)\bar{y}(x)]}{p_U(x)} \right\}^2 p_U(x) dx = \alpha^2,$$

где  $\alpha > 0$  задана, то параметр  $\lambda$  в (18) определится как

$$(21) \quad \lambda = \alpha \left[ \int_{R^n} (f(x) - c)^2 p_U(x) dx \right]^{-1/2}.$$

Соответствующая величина  $\delta_Y F(X)$ :

$$(22) \quad \delta_Y F(X) = \alpha \left[ \int_{R^n} (f(x) - c)^2 p_U(x) dx \right]^{1/2}.$$

Если функцию  $p_U(x)$  интерпретировать как плотность распределения вероятности случайного вектора  $U$ , то (22) и (19) можно переписать в виде

$$(23) \quad \delta_Y F(X) = \alpha \{\sigma^2[f(U)]\}^{1/2}, \quad c = E[f(U)],$$

где  $\sigma^2[f]$  – дисперсия  $f(U)$ .

Суммируем полученные результаты.

**Предложение 3.** Градиентное направление  $Y$ , определяемое векторным полем (13), которое максимизирует  $\delta_Y F(X)$  при ограничениях (14), (20), удовлетворяет уравнениям (18), (19). Соответствующее значение  $\delta_Y F(X)$  при этом направлении определяется (22) или (23).

Градиентное направление определено с точностью до функции  $p_U(x) > 0$ . Концентрируя  $p_U(x)$  в произвольной точке  $x_0$ :  $p_U(x) \rightarrow \delta(x - x_0)$ , можно уменьшить нелокальность поиска:  $\delta_Y F(X) \rightarrow 0$  в (23). Наоборот, увеличение нелокальности поиска достигается за счет концентрации плотности  $p_U(x)$  в точках максимума и минимума  $f$  (что приводит к максимизации дисперсии (23)).

Легко видеть близость приведенных конструкций с выводом неравенства Крамера–Рао [19]. В частности, интеграл в (20) можно ассоциировать с функцией информации Фишера и интерпретировать как меру информации, доставляемой случайным вектором  $U$  (который, в соответствии с замечанием, сделанным выше, характеризует нелокальность поиска).

## 5. Потенциальная функция

Приведенное понятие градиентного векторного поля пока неконструктивно, поскольку множество решений (18) бесконечно. Для описания множества решений этого уравнения воспользуемся возможностью представления финитного или исчезающего на бесконечности векторного поля  $p(x)\bar{y}(x)$  в виде

$$(24) \quad p(x)\bar{y}(x) = \nabla\varphi(x) + w(x),$$

$$(25) \quad \operatorname{div} w(x) = 0.$$

При дополнительном условии  $\varphi|_{\infty} = 0$  такое представление является однозначным. Учитывая, что  $\operatorname{div} \operatorname{grad} \varphi = \Delta\varphi$ , где  $\Delta$  – оператор Лапласа, получим из (18), (24), (25)

$$(26) \quad \Delta\varphi = -\lambda[f(x) - c]p_U(x).$$

Таким образом, потенциальная функция, характеризующая градиентное направление, должна удовлетворять уравнению Пуассона (в неограниченной среде). Его решение, стремящееся к нулю на бесконечности (что естественно допустить в силу свойств плотности  $p_U(x)$ ), единственно и имеет вид

$$(27) \quad \varphi(x) = -\lambda \int_{R^n} E(x, \zeta) f(\zeta) - c) p_U(\zeta) d\zeta,$$

где  $E(x, \zeta)$  – фундаментальное решение уравнения Лапласа.

Из (27) получим выражение для градиента потенциальной функции:

$$(28) \quad \nabla\varphi(x) = -\lambda \int_{R^n} \nabla_x E(x, \zeta) (f(\zeta) - c) p_U(\zeta) d\zeta$$

или

$$(29) \quad \nabla\varphi(x) = -\lambda E[f(U) - c] \nabla_x E(x, U).$$

Таким образом, в достаточно общих предположениях о функции  $f$ , затрагивающих, по существу, лишь ее интегральные свойства (имеется в виду законность операции дифференцирования интеграла по параметру (11)), было получено представление (24), (25), (28) для направления движения, обеспечивающее максимизацию производной по направлению функционала  $F(X)$  в экстремальной задаче (2), эквивалентной исходной задаче (1). Такое представление (также при общих и естественных предположениях) единственно. Поэтому вектор  $\nabla\varphi$  в (28) может рассматриваться как градиент исходной функции  $f$ , использующий в своем выражении лишь ее значения; ньютоновский потенциал  $\varphi$  (27) может интерпретироваться как "портрет", модель исходной функции  $f$ . Эта модель определяется случайным вектором  $U$  или, в соответствии с (19), (23), средним значением  $f$  по распределению  $U$  и меняется при изменении наших знаний о минимальной величине  $f$ :  $\varphi = \varphi(x; c)$ .

Изложение этого раздела в основном следует [16]. Отметим в заключение, что потенциальная функция вида (27) для поиска глобального экстремума на эвристической основе использовалась в [20].

## 6. Численные методы

В этом разделе кратко рассмотрим численные методы, использующие потенциал  $\Phi$  (подробнее см. [12, 13]).

Методы первого порядка могут быть записаны в виде

$$(30) \quad X_{N+1} = X_N - \epsilon_N \nabla \Phi_N(X_N),$$

где  $\nabla \Phi_N$  определен в (28), (29) при  $c = c_N$ ,  $p_U = p_{U_N}(x)$ . Вследствие представления  $\nabla \Phi$  в виде математического ожидания (29), в (30) естественно воспользоваться вероятностными итеративными методами, использующими реализации случайных векторов [7, 21]. В этом случае длина шага должна удовлетворять известным условиям:

$$(31) \quad \sum_{N=0}^{\infty} \epsilon_N = \infty, \quad \sum_{N=0}^{\infty} \epsilon_N^2 < \infty, \quad \epsilon_N > 0.$$

Методы второго порядка имеют вид

$$(32) \quad X_{N+1} = X_N - [\nabla^2 \Phi_N(X_N)]^{-1} \nabla \Phi_N(X_N).$$

Для их реализации необходима оценка  $[\nabla^2 \Phi_N]^{-1}$ . При довольно общих предположениях, использующих понятие сингулярного интеграла, можно показать [11], что

$$(33) \quad \nabla^2 \Phi(x) = p(x)(\hat{f}(x)/n)I + E[v(U, x)]I - nE[v(U, x)\theta\theta^T],$$

где  $I$  – единичная матрица,  $\hat{f}(x) = f(x) - c$ ,  $\theta = (U - x)/\|U - x\|$ ;

$$v(U, x) = \chi_{R^n \setminus S_p}(U) \frac{\hat{f}(U)}{\omega_n \|U - x\|^n} + \chi_{S_p}(U) \frac{\hat{f}(U) - \hat{f}(x) \frac{p_U(x)}{p_U(U)}}{\omega_n \|U - x\|^n},$$

$$\chi_{S_p}(U) = \begin{cases} 1, & \|U - x\| \leq \rho \\ 0, & \|U - x\| > \rho \end{cases}, \quad \rho > 0,$$

$\omega_n$  – площадь поверхности единичной сферы в  $R^n$ .

Можно показать [12, 15], что структура матрицы  $\nabla^2 \Phi$  содержит операции растяжения и отражения (соответственно в перспективном и неперспективном направлениях), которые являются основой современных методов оптимизации [22] и позволяют существенно упростить операцию обращения  $\nabla^2 \Phi$ . Таким образом, эти операции с необходимостью входят в структуру алгоритмов ньютоновского типа (32), не использующих дифференциальные характеристики исходной функции  $f$ .

Анализ сходимости методов (30), (32) был проделан в [17], где было показано, что в качестве функции Ляпунова можно использовать потенциал  $\Phi$ .

Отметим в заключение, что к методам второго порядка можно прийти, если вместо  $\nabla^2 \Phi$  использовать вариацию второго порядка функционала  $F$  в (2) [11].

## 7. Методы возможных направлений

Как было отмечено в разделе 2, вместо градиентного направления  $Y$  (максимизирующего  $\delta_Y F(X)$ ), можно использовать такие направления, для которых  $\delta_Y F(X) \leq 0$ . В этом случае получим из (15) и (24), (25), что

$$(34) \quad \int_{R^n} (f(x) - c) \Delta \Phi(x) dx \geq 0.$$



Из (34) следует

$$(35) \quad \Delta\varphi(x) \geq 0, \quad x \in \Omega_n = \{x | f(x) > c\},$$

$$\Delta\varphi(x) \leq 0, \quad x \in \Omega_n = \{x | f(x) \leq c\}.$$

Следовательно, в неперспективной области  $\Omega_n$  потенциальная функция  $\varphi$  является субгармонической, в то время как в перспективной области  $\Omega_n$  – супергармонической [23].

В соответствии с классической теоремой Рисса любая субгармоническая функция может быть представлена в виде суммы гармонической и потенциальной функций [23].

Наличие в структуре субгармонической функции потенциала  $\varphi$  показывает, что в структуре векторного поля  $\bar{y}(x)$ , характеризующего направление движения, должна присутствовать составляющая  $\nabla_x E(x, \zeta) = \omega_n \|x - \zeta\|^{n-2} (x - \zeta)$ . Следовательно, ядро  $\nabla_x E(x, \zeta)$  образует структурную (универсальную) составляющую процедур нелокального поиска, использующих как градиентные, так и в общем случае возможные направления.

Проанализируем особенности потенциальной функции более подробно. Разобьем  $\varphi$  на две составляющие:

$$(36) \quad \varphi(x) = \int_{\Omega_n} E(x, \zeta) \tilde{f}(\zeta) d\zeta + \int_{\Omega_n} E(x, \zeta) \tilde{f}(\zeta) d\zeta = \varphi_n(x) + \varphi_n(x),$$

где  $\tilde{f}(\zeta) = (f(\zeta) - c)p_U(\zeta)$ .

Очевидно, что функция  $\varphi_n(x)$  является в  $R^n$  супергармонической (гармонической в  $\Omega_n$ ), а функция  $\varphi_n(x)$  – субгармонической в  $R^n$  (в  $\Omega_n$  – гармонической). В результате векторное поле  $\bar{y}(x)$  представится в виде

$$(37) \quad p(x)\bar{y}(x) = \nabla\varphi_n(x) + \nabla\varphi_n(x).$$

Если поисковая точка  $x$  оказывается в неперспективном множестве  $\Omega_n$ , то  $\delta_Y F(X) = 0$  для  $\nabla\varphi_n$ , так как  $\varphi_n$  гармонична в  $\Omega_n$  ( $\Delta\varphi_n(x) = 0, x \in \Omega_n$ ). Поэтому ограничимся здесь (т.е. в  $\Omega_n$ ) составляющей направления движения  $\nabla\varphi_n$ , субгармонична. В силу принципа максимума для субгармонических функций [23, 24] максимальное значение  $\varphi_n$ , равное нулю (так как  $f(x) - c \geq 0$  и  $E(x, \zeta) < 0$  для  $x \in \Omega_n$ ), не достигается внутри области  $\Omega_n$ . Таким образом, поисковые движения, осуществляемые в направлении  $\nabla\varphi_n$  (т.е. соответствующие максимизации  $\varphi_n(x)$ ), приводят к "выталкиванию" поисковой точки из неперспективной области  $\Omega_n$ . Так как максимизация  $\varphi_n$  эквивалентна нахождению таких точек  $x$ , для которых  $\varphi_n = 0$ , то плотность  $p_U(x)$  целесообразно концентрировать именно в этих точках, что соответствует решению уравнения  $E(x_0, \zeta)(f(\zeta) - c) = 0$ , где  $x_0$  – реализация текущего случайного вектора  $X_0$  (т.е. в окрестности локального минимума  $f$ ). Тогда член  $\|x_0 - \zeta\|^{n-2}$  в знаменателе  $E(x_0, \zeta)$  и будет обеспечивать выталкивание поисковой точки из окрестности локальных минимумов. В этом идея туннельных алгоритмов [25].

В заключение отметим, что использование ньютонова потенциала (27) является, очевидно, не единственным способом обеспечить неравенство  $\delta_Y F(X) \leq 0$ . В частности, волновой потенциал  $\psi(x)$ , который является решением волнового уравнения [24, 26]

$$(38) \quad \Delta\psi(x) + \omega^2\psi(x) = (f(x) - c)p_U(x),$$

также может служить как основа для развития нелокальных методов поиска [11, 17].

## 8. Граничные условия

На этапе постановки задачи или обработки текущей информации в процессе решения задачи область поиска может сужаться, т.е. в общем случае нужно решать задачу (7), а не (2). В этом случае производная по направлению будет иметь вид

$$(39) \quad \delta_Y F(X) = \int_{\Omega} (\nabla f(x), p(x) \bar{y}(x)) dx.$$

Используя формулу Стокса, получим

$$\delta_Y F(X) = \int_{\partial\Omega} (f(x) - c)(p(x) \bar{y}(x), n(x)) ds - \int_{\Omega} (f(x) - c) \operatorname{div}[p(x) \bar{y}(x)] dx,$$

где  $n(x)$  – единичная нормаль к границе  $\partial\Omega$ . Отсюда и из представления (24), (25) будем иметь

$$(40) \quad \delta_Y F(X) = \int_{\partial\Omega} (f - c) \left( \frac{\partial \varphi}{\partial n} + w_n \right) ds - \int_{\Omega} (f - c) \Delta \varphi dx,$$

где  $\partial \varphi / \partial n = (\nabla \varphi(x), n(x))$ ,  $w_n = (w(x), n(x))$ .

Условие улучшения  $\delta_Y F(X) \leq 0$  для (40) приводит в этом случае к краевой задаче Неймана [24] (при  $w_n = 0$ ):

$$(41) \quad \begin{aligned} \Delta \varphi(x) &= (f(x) - c) p_U(x), \quad x \in \Omega, \\ \partial \varphi(x) \partial n &= -(f(x) - c) p_V(x), \quad x \in \partial\Omega, \end{aligned}$$

где  $p_U(x), p_V(x)$  – некоторые неотрицательные функции, которые могут интерпретироваться как плотности распределения вероятностей случайных векторов  $U$  и  $V$ , выбираемых на основе текущей информации в процессе решения задачи.

Таким образом, краевая задача Неймана естественно возникает в процессе построения потенциального поля для нелокального поиска.

Можно также рассмотреть и краевую задачу Дирихле

$$(42) \quad \begin{aligned} \Delta \varphi(x) &= (f(x) - c) p_U(x), \quad x \in \Omega, \\ \varphi(x) &= -(f(x) - c) p_V(x), \quad x \in \partial\Omega. \end{aligned}$$

Однако в этом случае граничное условие постулируется (несколько искусственно), в отличие от (41), где оно выводится из (40).

Решение краевых задач (41), (42) может быть получено с помощью функции Грина. Таким образом, в общем случае универсальная составляющая нелокального поиска, определяемая ядром в (27), заменяется функцией Грина  $G(x, \zeta)$ .

## 9. Учет дифференциальных свойств оптимизируемой функции

В предыдущих построениях информация о дифференциальных свойствах  $f$  не использовалась. Если такая информация имеется, то для ее учета перепишем (30) в виде

$$(43) \quad \nabla \varphi(x) = -\lambda \int_{R^n} \nabla_{\zeta} E(x, \zeta) (f(\zeta) - c) p_U(\zeta) d\zeta$$

и, применяя к (43) формулу Грина, получим

$$\nabla \varphi(x) = -\lambda \int_{R^n} \nabla [(f(\zeta) - c) p_U(\zeta)] E(x, \zeta) d\zeta$$

или, окончательно,

$$(44) \quad \nabla \varphi(x) = -\lambda \int_{R^n} E(x, \zeta) \nabla f(\zeta) p_U(\zeta) d\zeta - \lambda \int_{R^n} E(x, \zeta) (f(\zeta) - c) \nabla p_U(\zeta) p_U(\zeta) d\zeta.$$

Таким образом, локальная информация о  $\nabla f$  (первый член в (44)) добавляется ко второму "нелокальному" члену в (44).

## 10. Структурные составляющие нелокального поиска

Проанализируем последовательность предположений и построений, приведших к определению потенциала и алгоритмов его минимизации. Прежде всего отметим, что переход от исходной задачи к рандомизированной является в рассматриваемом случае эквивалентным (в смысле асимптотического решения) и лишь расширяет возможности построения процедур нелокального поиска. Последующие построения делались также при естественных и неограничивающих предположениях (типа существования интеграла или законности операции дифференцирования интеграла по параметру). Поэтому можно говорить о структурных (универсальных) составляющих, с необходимостью присутствующих в процедурах нелокального поиска.

Действительно, анализ построений (1) – (30), приведших к построению потенциала, показывает, что ядро  $E(x, \zeta)$  (или в общем случае функция Грина  $G(x, \zeta)$ ) является такой универсальной составляющей нелокального поиска, не зависящей ни от способа рандомизации, ни от выбора конкретной целевой функции  $f$ . Анализ производных первого порядка этого ядра показывает, что в нелокальном поиске с необходимостью присутствует неустойчивая составляющая в неперспективной области, а анализ производных второго порядка приводит (также с необходимостью) к операциям растяжения – отражения пространства в нелокальном поиске.

В заключение кратко остановимся на роли векторного поля  $w(x)$  в нелокальном поиске. Подставляя (24), (25) в (18), (22) и предполагая, что  $\Delta\phi(x) = 0$ , получим, что  $\delta_\gamma F(X) = 0$ . Таким образом, движение, которое определяется векторным полем  $w(x)$ , не меняет среднего значения функции и необходимо для накопления знаний о ней. Поэтому неоднозначность представления (24), (25), связанная с выбором  $w(x)$ , может рассматриваться как своего рода "плата" за отсутствие априорной информации о наблюдаемой функции.

## 11. Заключение

В настоящей работе основное внимание было уделено теоретическим аспектам рассматриваемого направления. Численные методы подробно рассмотрены в [12, 13]. Следует отметить преемственность этих алгоритмов (на уровне основных операций или структурных составляющих) с известными детерминированными аналогами (методы первого и второго порядка, сопряженных направлений, растяжения пространства, эллипсоидов и др.). При этом в алгоритмах, основанных на использовании потенциальной функции, вычисляются лишь значения исходной целевой функции.

Наконец, в заключение отметим, что в этой работе использовался потенциал, зависящий от интегральных свойств оптимизируемой функции  $\phi = \phi(x; c)$ .  $c = \int f(x)p_U(x)dx$ . Возможно построение потенциальной функции, зависящей от усредненного решения ("центра" аппроксимации):  $\phi = \phi(x, \bar{c})$ ;  $\bar{x} = \int xp_U(x)dx$ . В этом случае соответствующие методы будут близки к методам локальной аппроксимации [27] или проекционным и ядерным оценкам плотности [28] и кратко рассмотрены в [29].

## СПИСОК ЛИТЕРАТУРЫ

1. Fromovitz S. Nonlinear programming with randomization // Management Sci. 1965. V. 11. № 9. P. 831–846.

2. Каплинский А.И., Пропой А.И. О стохастическом подходе к задачам нелинейного программирования // *Авт.* 1970. № 3. С. 122–133.
3. Ермольев Ю.М. Об одной общей задаче стохастического программирования // *Кибернетика*. 1971. № 3. С. 47–50.
4. Юдин Д.Б. Математические методы управления в условиях неопределенности. М.: Советское радио, 1974.
5. Каплинский А.И., Позняк А.С., Пропой А.И. Условия оптимальности для некоторых задач стохастического программирования // *Авт.* 1971. № 8. С. 51–60.
6. Каплинский А.И., Позняк А.С., Пропой А.И. О некоторых методах решения задач стохастического программирования // *Авт.* 1971. № 10. С. 87–94.
7. Цыпкин Я.З. Адаптация и обучение в автоматических системах. М.: Наука, 1968.
8. Айзерман М.А., Браверман Э.М., Розоноэр Л.И. Методы потенциальных функций в теории обучения машин. М.: Наука, 1970.
9. Пропой А.И., Пухликов А.В. Основания математической теории систем. М.: ВНИИСИ, 1990.
10. Каплинский А.И., Лимарев Е.Е., Чернышева Г.Д. Построение рандомизированных алгоритмов оптимизации // *Проблемы случайного поиска*. 1980. Вып. 8. С. 63–91.
11. Каплинский А.И., Пропой А.И. Вариационный подход к построению алгоритмов нелокальной оптимизации. М.: ВНИИСИ, 1986.
12. Каплинский А.И., Пропой А.И. Структурные составляющие методов нелокального поиска, использующих теорию потенциала. М.: ВНИИСИ, 1990.
13. Каплинский А.И., Пропой А.И. Конструирование вычислительных алгоритмов нелокального поиска, использующих теорию потенциала. М.: ВНИИСИ, 1990.
14. Гиль А.Б., Каплинский А.И., Пропой А.И. Построение вычислительных схем нелокальной оптимизации на основе теории потенциала // *Статистические методы и модели*. Сб. трудов. М.: ВНИИСИ. 1987. Вып. 1. С. 11–23.
15. Гиль А.Б., Каплинский А.И., Пропой А.И. Об использовании теории потенциала в вариационном подходе к построению алгоритмов оптимизации, использующих операции отражения и растяжения // *Модели и методы оптимизации*. Сб. трудов. М.: ВНИИСИ. 1987. Вып. 11. С. 72–78.
16. Каплинский А.И., Пропой А.И. О градиентной основе негладкой оптимизации, использующей теорию потенциала // *Задачи и методы оптимизационного моделирования*. Сб. трудов. М.: ВНИИСИ, 1988. Вып. 13. С. 11–16.
17. Каплинский А.И., Песин А.М., Пропой А.И. О методах нелокального поиска // *Модели и методы оптимизации*. Сб. трудов. М.: ВНИИСИ, 1991. Вып. 11. С. 35–45.
18. Егоров Ю.В. Линейные дифференциальные уравнения главного типа. М.: Наука, 1984.
19. Закс Ш. Теория статистических выводов. М.: Мир, 1975.
20. Красовский А.А. Непрерывные алгоритмы и стохастическая динамика поиска экстремума // *Авт.* 1991. № 4. С. 55–65.
21. Невельсон М.Б., Хасьминский Р.З. Стохастическая аппроксимация и рекуррентное оценивание. М.: Наука, 1972.
22. Шор Н.З. Методы минимизации недифференцируемых функций и их приложения. Киев: Наукова думка, 1979.
23. Хейман Х., Кеннеди П. Субгармонические функции. М.: Мир, 1980.
24. Тихонов А.Н., Самарский А.А. Уравнения математической физики. М.: Наука, 1972.
25. Levi A.V., Montalvo A. The tunneling algorithm for the global minimization of functions // *SIAM J. Sci. Comp.* 1985. V. 6. № 1. P. 15–29.
26. Санчес-Паленсия Э. Неоднородные среды и теория колебаний. М.: Мир, 1984.
27. Катковник В.Я. Линейные оценки и стохастические задачи оптимизации. М.: Наука, 1976.
28. Деврой Л., Дьерфи Л. Непараметрическое оценивание плотности. М.: Мир, 1988.
29. Пропой А.И. О некоторых принципах нелокального поиска // *Оптимизация и управление в сложных системах*. Сб. трудов М.: ВНИИСИ, 1986. Вып. 4. С. 67–80.

Поступила в редакцию 25.06.92.